Revista Argentina de Ciencias del Comportamiento, 2010, Vol. 2, N°2, 52-59

ISSN 1852-4206
www.psyche.unc.edu.ar/racc



Revista Argentina de Ciencias del Comportamiento (RACC)

Vista: Un Software para la Enseñanza de la Estadística y la Psicometría

Ledesma, Ruben D.[*a], Valero-Mora, Pedro [b] y Molina, Jesús G. [b]

[a] Universidad Nacional de Mar del Plata (CONICET - UNMDP), Mar del Plata, Argentina
[b] Universidad de Valencia, Valencia, España

Artículo Metodológico

Resumen

La enseñanza moderna del análisis de datos supone el uso de algún tipo de software estadístico, existiendo actualmente una amplia variedad de opciones para este fin. Este trabajo analiza e ilustra las potencialidades del sistema ViSta "The Visual Statistics System" en la formación estadística. Como desarrolladores actuales de este programa, entendemos que las siguientes características resultan de interés para fines educativos: a) acceso libre: es un programa gratuito y disponible on-line; b) flexibilidad: es un sistema abierto, extensible y adaptable a diferentes usuarios y necesidades; c) simplicidad de la interfaz: posee una interfaz gráfica de usuario simple e intuitiva que facilita la gestión, transformación y análisis de datos; y d) capacidad gráfica: implementa métodos de visualización que permiten realizar demostraciones de conceptos y técnicas de análisis. Estas capacidades se ilustran mediante una aplicación específica que hemos desarrollado en ViSta para la enseñanza de conceptos de Psicometría Clásica. Se discuten las limitaciones y perspectivas de desarrollo del sistema.

Palabras claves:
Enseñanza Estadística, Análisis de Datos, Gráficos Estadísticos, Psicometría, Software Estadístico

Abstract

Vista: A Software for Teaching Statistics and Psychometrics. Nowadays, teaching statistics is almost undisputedly always done using some sort of software for carrying out the computations. Indeed, the range of choices available for that purpose is wide and with a lot of variety. The present work describes the capabilities of one of the systems currently available for teaching of statistics: ViSta, the Visual Statistics System. As ViSta's developers, we wish to point out the following features as specially appealing for using it at the classroom: a) price and availability, it is free and can be downloaded on-line; b) flexibility, it is an open system, extensible and adaptable for different users and needs; c) simplicity, it has a simple and intuitive user interface that facilitates the management, transformation and data analysis; and d) its graphical power, as it implements visualization methods that allow carrying out demonstrations of concepts and data analysis techniques. These capabilities are illustrated using a specific application to the teaching of psychometrics. The limitations and perspectives of development of the system are finally discussed.

Key Words:
Teaching of Statistics, Data Analysis, Statistical Graphics, Psychometrics, Statistical Software

Recibido el 18 de Mayo de 2010; Recibido la revisión el 14 de Junio de 2010; Aceptado el 14 de Junio de 2010

1. Introducción

En la actualidad, la enseñanza del análisis de datos en carreras de ciencias supone el uso de algún tipo de software estadístico, existiendo una amplia variedad de recursos informáticos de diferente alcance y naturaleza que pueden ser utilizados para este fin. Frente a la variedad de opciones disponibles, es necesario evaluar y seleccionar adecuadamente el software en función de las necesidades y objetivos de enseñanza, teniendo en cuenta, a la vez, las posibilidades reales de acceso o adquisición de los productos y licencias. Para un análisis de las posibles ventajas y desventajas comparativas de las herramientas informáticas en el contexto de la enseñanza, comenzaremos por señalar algunas distinciones básicas relacionadas con las características del software.

2. Software de propósitos múltiples versus programas especializados

Una primera cuestión tiene que ver con el

[*] Enviar correspondencia a: Dr. Ledesma, Ruben D.
E-mail: rdledesma@conicet.gov.ar

alcance de los programas en cuanto al rango o variedad de métodos estadísticos que incluyen. Aquí, encontramos desde los paquetes o sistemas de propósito múltiple, tipo SPSS (Spss Inc., 2003), Statistica (StatSoft Inc., 2005) o Stata (StataCorp LP, 2009), hasta los programas especializados en la aplicación de una sola técnica estadística. Los primeros incluyen herramientas de gestión, transformación, análisis y presentación de resultados, incorporando un amplio abanico de técnicas estadísticas básicas y avanzadas. Estos sistemas, sobre todo SPSS, son ampliamente utilizados en ciencias del comportamiento. Por otro lado, existen programas especializados en ciertos métodos específicos, los cuales son capaces de realizar un número limitado de análisis, aunque generalmente de forma más potente y flexible que el software convencional. Ejemplos pueden ser los programas Lisrel o AMOS para modelos de ecuaciones estructurales, HLM para modelos lineales jerárquicos, Arc para modelos de regresión, etc.

Los sistemas de propósito múltiple pueden ser más apropiados para cursos generales de análisis de datos y, también, para prácticas de investigación que implican carga, edición, gestión de bases de datos, etc. Al contrario, si el curso trata específicamente sobre un tema o conjunto de técnicas puede ser más funcional utilizar un sistema estadístico especializado. En ocasiones, esta es la única opción, pues ciertos métodos no están disponibles en los programas de propósito múltiple o no se incluyen con el grado de flexibilidad y sofisticación requeridas. Por ejemplo, SPSS no permite aplicar técnicas de remuestreo de datos o modelos de ecuaciones estructurales, en estos casos necesitaríamos un software específico.

3. Software de usuario y lenguajes de programación (R, LispStat, etc.)

También es conveniente distinguir entre el software estadístico orientado principalmente a realizar aplicaciones a nivel de usuario y los lenguajes de programación estadística, más apropiados para la innovación y el desarrollo de métodos en estadística computacional. *R* (*R* Development Core Team, 2002) y LispStat (Tierney, 1990) son dos exponentes de este tipo. En ambos casos se trata de poderosos lenguajes de programación estadística, pero generalmente requieren un cierto grado de experticia, más propio de un usuario avanzado que de un estudiante. Si bien existen interfaces gráficas de usuario para estos lenguajes, aún no terminan de ser del todo amigable para quienes se inician en la materia. Está claro, no obstante, que R se encuentra en plena expansión y es de esperar que en el futuro próximo avancen sensiblemente los desarrollos en materia de interfaz gráfica de usuario.

Por otro lado, también es verdad que la interacción con el lenguaje puede mejorar la comprensión y proporcionar un conocimiento más detallado y en profundidad de los métodos, lo que puede ser más provechoso para estudiantes de ciencias que poseen formación matemática. Sin embargo, si se trata de usuarios menos experimentados, estos lenguajes pueden resultar tediosos y quitar dinamismo a la enseñanza estadística. Si el objetivo del curso es brindar elementos para que el estudiante pueda aprender a codificar, cargar y analizar datos con fines de investigación - que puede ser el caso de las carreras en ciencias humanas y sociales -, resulta necesario disponer de sistemas más amigables y flexibles que permitan realizar estas tareas de modo sencillo, cuestión que en los entornos como R aún resulta difícil de lograr. En todo caso, es importante reconocer que existe una cierta 'tensión', difícil de resolver, entre la simplicidad de uso y la flexibilidad de las herramientas. En este sentido, el docente deberá ponderar diferentes aspectos, desde los conocimientos informáticos previos de sus estudiantes hasta los objetivos y el contexto global de enseñanza.

4. Accesibilidad del software

Otro elemento clave y determinante para la elección de un software es su accesibilidad, es decir, las posibilidades de ser adquirido y las condiciones bajo las cuales se puede distribuir. Cabe mencionar que algunos paquetes estadísticos profesionales son prácticamente inaccesibles para docentes y estudiantes en nuestro contexto. Los costos asociados a su adquisición resultan difíciles de cubrir en el ámbito público, donde los recursos

destinados a la compra de software son realmente escasos. Frente a estas limitaciones, tenemos la opción de programas completamente gratuitos que pueden obtenerse on-line y utilizarse perfectamente en el ámbito educativo. Entre estas herramientas encontramos sistemas de propósito múltiple, como EpiInfo (CDC, 2000), WinIDAMS (UNESCO, 2002), ViSta (Young, 1996) y OS (Miller, 2003), como así también una gran cantidad de programas especializados.

Podemos afirmar que actualmente el software gratuito se plantea como una alternativa viable frente a las restricciones de acceso que para muchos sectores plantea el software de origen comercial. Como añadido, las herramientas gratuitas pueden llegar a superar en ciertos aspectos a los sistemas comerciales. Por ejemplo, si consideramos comparativamente las capacidades en materia de gráficos estadísticos, es evidente que el programa ViSta ofrece al usuario técnicas más sofisticadas y dinámicas que las que puede encontrar en sistemas como SPSS (Young, Valero-Mora & Friendly, 2006). En cualquier caso, lo importante aquí es destacar que las cuestiones comparativas en cuanto a las prestaciones y calidad del software están supeditadas en la práctica a la accesibilidad. Una revisión más detallada del software estadístico gratuito y sus posibilidades puede encontrarse en Ledesma (2004).

5. ViSta "The Visual Statistics System" en la Enseñanza de la Estadística

Considerando los puntos anterior, creemos que ViSta reúne características de interés para la docencia. Se trata de un sistema estadístico de propósito múltiple, gratuito y basado en un lenguaje de programación abierto, el LispStat. ViSta fue creado y desarrollado por el profesor Forrest W. Young en la Universidad de Carolina del Norte en Chapel Hill (Young, 1996), siendo el resultado de varios años de investigación en el área de la visualización de datos y de su experiencia como profesor de estadística en Psicología. Si bien se trata de un programa diseñado originalmente como herramienta para la investigación y desarrollo de métodos de visualización estadística multivariada, actualmente posee características que lo definen como un sistema general, pues integra una variedad de herramientas de procesamiento y análisis de datos. En particular, ViSta destaca entre otros programas gratuitos por sus técnicas gráficas innovadoras, basadas en métodos dinámicos e interactivos.

Actualmente ViSta se presenta como un proyecto de software abierto en el que participan desarrolladores y usuarios interesados en el programa. En nuestro caso, hemos utilizado ViSta como investigadores interesados en la creación de métodos de visualización (Ledesma, Molina, Valero-Mora y Poó, 2007), pero también como docentes a cargo de cursos de análisis de datos en ciencias del comportamiento. Nuestra experiencia a través de varios años de docencia en el área sugiere que ViSta puede ser una herramienta útil para mejorar la calidad de la enseñanza y el aprendizaje de la estadística aplicada. A continuación detallaremos las principales características técnicas que hacen de ViSta un programa potencialmente útil para docentes y estudiantes.

a) Accesibilidad: ViSta es un programa completamente gratuito disponible en Internet. La versión más conocida y extendida del programa es WinViSta 6.4, que puede obtenerse en la dirección: http://forrest.psych.unc.edu/research/index.html. Una versión más reciente, con mejores capacidades, puede descargarse desde el sitio: www.uv.es/visualstats/Book/. En síntesis, para obtener ViSta el docente y los estudiantes solo necesitan una conexión a Internet. Luego de la descarga, la instalación es automática y se realiza sin ningún tipo de restricciones. Esto es una ventaja importante en el contexto de la enseñanza.

b) Flexibilidad: ViSta es un sistema que podríamos denominar 'multi-usuario'. Puede adecuarse a un amplio rango de necesidades y niveles de experticia. Como hemos mencionado, este programa integra en un mismo entorno una variedad de recursos para procesar y analizar datos mediante diferentes modelos estadísticos, desde técnicas básicas hasta modelos multivariados. De esta manera se proporciona un abanico de posibilidades para diferentes problemas y tipos de datos. Además, como ViSta está basado en LispStat, puede explotarse la versatilidad de este lenguaje para mejorar, extender o adecuar el

programa a las propias necesidades. Así, el docente puede aprovechar la variedad de técnicas existentes para enseñar diferentes modelos y técnicas, pero también, dependiendo de su conocimiento en programación, puede adaptar las herramientas ya disponibles o bien crear sus propios recursos y materiales didácticos.

c) Simplicidad: ViSta fue concebido en el contexto de la enseñanza estadística en Psicología. Dispone de una interfaz de usuario completamente gráfica que permite a los estudiantes una interacción simple e intuitiva. Esta interfaz gráfica tiene un formato convencional de menús, cuadros de diálogos y otras técnicas integradas en un escritorio de trabajo. De este modo, el usuario familiarizado con interfaces gráficas puede realizar tareas de gestión, transformación, análisis y presentación de resultados de un modo relativamente sencillo. Así, se simplifica y facilita la tarea de quienes se inician en el aprendizaje del análisis estadístico de datos, permitiendo además centrar el foco de la tarea docente en cuestiones sustantivas, prescindiendo de mayores explicaciones en lo referente al funcionamiento del programa.

d) Potencia gráfica: Varios autores han señalado la importancia y el poder de los gráficos no solo como medio de análisis, sino también como recurso educativo (e.g., Young, Valero-Mora y Friendly, 2006). Entre los gráficos estadísticos, resultan de especial interés los gráficos dinámicos e interactivos, que pueden usarse en la docencia para ilustrar conceptos estadísticos y también como una herramienta para facilitar la exploración y manipulación directa de los datos. Esta interacción directa mejora no solo la potencia exploratoria del análisis, sino también la comprensión de los datos y conceptos. Como ya hemos comentado, ViSta destaca por sus métodos de visualización dinámicos e interactivos, incluyendo actualmente más de 40 'spreadplots' (gráficos extendidos) que pueden usarse para diferentes propósitos y en apoyo a las salidas numéricas en formato texto. Los spreadplots son un enfoque gráfico novedoso que aprovecha las capacidades de LispStat en materia de gráficos dinámicos (Young, Valero-Mora, Faldowski y Bann, 2003). Se trata de un método que permite relacionar y coordinar conjuntos de gráficos dinámicos e interactivos, presentados simultáneamente en una misma visualización. Los spreadplots en ViSta se utilizan para visualizar simulaciones, transformaciones de variables, diferentes tipos de datos, y resultados de modelos estadísticos.

Creemos que las características mencionadas posicionan a ViSta como un software accesible y atractivo para fines educativos. A continuación ilustramos con mayor detalle las capacidades de ViSta mediante una aplicación específica que hemos desarrollado para la enseñanza de la Psicometría.

6. Ejemplo de Aplicación en Psicometría

El módulo ViSta-CITA "Classical Item and Test Analysis" (Ledesma y Molina, 2009) es un plug-in, o módulo de análisis integrable a ViSta. Este programa proporciona opciones de análisis que habitualmente encontraríamos entre los contenidos de un curso básico de Psicometría, incluyendo análisis de ítems, fiabilidad, análisis factorial, errores de medida, normalización de puntajes, etc. El programa aún no incluye métodos de Teoría de Respuesta al Ítem (TRI), pero permite cubrir prácticamente todos los análisis de la Teoría Clásica de los Test (TCT). Insistiendo en las capacidades de ViSta para su utilización en el ámbito educativo, podemos destacar las siguientes características específicas de este módulo de análisis psicométrico:

1) Accesibilidad. ViSta-CITA proporciona una opción gratuita frente a otros programas de análisis que pueden usarse con fines psicométricos. El programa ya se encuentra integrado en la versión 7 de ViSta. Para la versión 6.4, existe un plug-in que se descarga desde Internet sin restricciones (www.mdp.edu.ar/psicologia/vista/). Luego de la descarga, el programa se podrá instalar de forma automática dentro de ViSta y usarse dentro de este entorno, aprovechando así el resto de las capacidades de análisis del programa.

2) *Flexibilidad.* El módulo ViSta-CITA ofrece una serie de métodos para cubrir varios tipos de análisis propios de la TCT. Se incluyen métodos de análisis de ítems, diferentes opciones de análisis de fiabilidad, corrección y estandarización de

puntajes, y análisis factorial exploratorio, entre otras opciones. De este modo, el docente dispone de todos las herramientas necesarias para ilustrar y ejemplificar aplicaciones en TCT. Además, el módulo está escrito en lenguaje LispStat, de modo que puede ser modificado, mejorado o ampliado en sus capacidades si el docente o especialista domina ciertos rudimentos de este lenguaje.

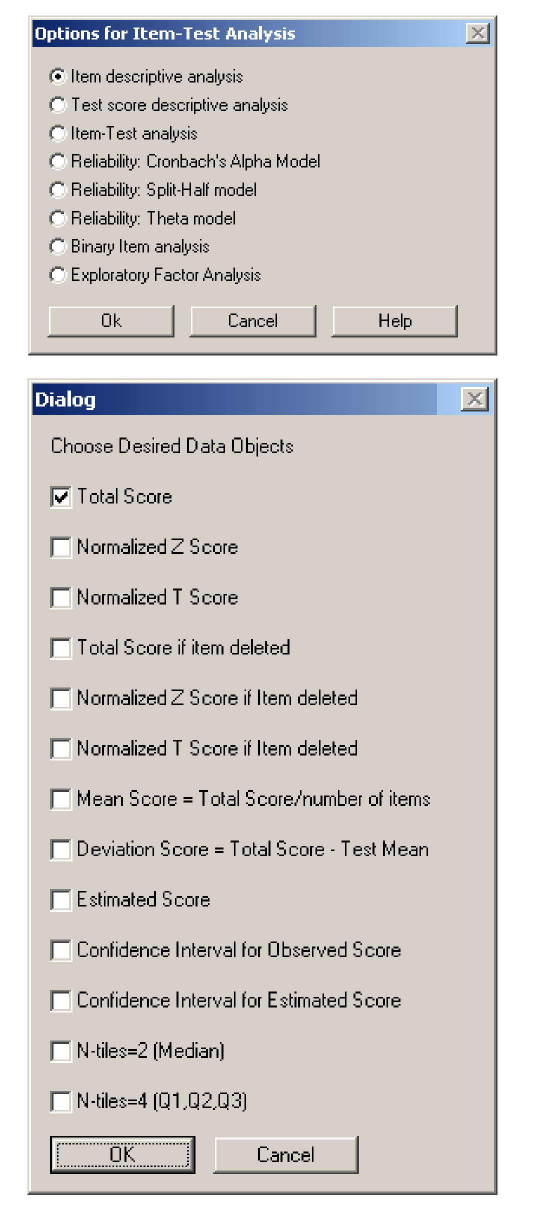
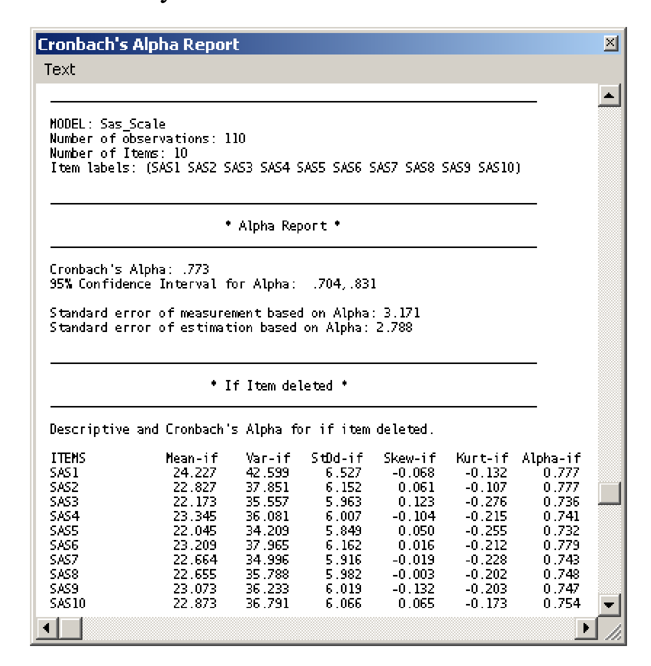


Figura 1. Imagen general de la interfaz de ViSta y cuadros de diálogo del programa ViSta-CITA.

3) Simplicidad. ViSta-CITA se integra completamente a la interfaz gráfica de ViSta, puede ejecutarse mediante menús, cuadros de diálogos, etc. La Figura 1 muestra el entorno de trabajo del programa ViSta y ejemplos de cuadros de diálogo para interactuar con ViSta-CITA. El usuario no necesita conocer el lenguaje LispStat para interactuar con el programa, lo cual puede ser beneficioso si se trabaja con estudiantes que solo necesitan una formación aplicada en TCT (e.g, estudiantes de carreras de Psicología, Educación, Ciencias de la Salud, etc.). Esto vale tanto para la carga y edición de datos, como para la ejecución del análisis y la obtención de salidas de resultados.


Cronbach's Alpha Report
Text

MODEL: Sas_Scale
Number of observations: 110
Number of Items: 10
Item labels: (SAS1 SAS2 SAS3 SAS4 SAS5 SAS6 SAS7 SAS8 SAS9 SAS10)

* Alpha Report *

Cronbach's Alpha: .773
95% Confidence Interval for Alpha: .704,.831

Standard error of measurement based on Alpha: 3.171
Standard error of estimation based on Alpha: 2.788

* If Item deleted *

Descriptive and Cronbach's Alpha for if item deleted.

ITEMS	Mean-if	Var-if	StDd-if	Skew-if	Kurt-if	Alpha-if
SAS1	24.227	42.599	6.527	-0.068	-0.132	0.777
SAS2	22.827	37.851	6.152	0.061	-0.107	0.777
SAS3	22.173	35.557	5.963	0.123	-0.276	0.736
SAS4	23.345	36.081	6.007	-0.104	-0.215	0.741
SAS5	22.045	34.209	5.849	0.050	-0.255	0.732
SAS6	23.209	37.965	6.162	0.016	-0.212	0.779
SAS7	22.664	34.996	5.916	-0.019	-0.228	0.743
SAS8	22.655	35.788	5.982	-0.003	-0.202	0.748
SAS9	23.073	36.233	6.019	-0.132	-0.203	0.747
SAS10	22.873	36.791	6.066	0.065	-0.173	0.754

Figura 2. Imagen de la salida numérica del programa ViSta-CITA

4) Potencia gráfica. ViSta-CITA ofrece salidas numéricas convencionales para cada análisis (ver figura 2), pero también complementa estas salidas con "spreadplots" (gráficos extendidos) especializados. Las figuras 3 y 4 ilustran el concepto de los spreadplots, aplicados en este caso a la visualización de resultados de análisis psicométricos. El primer caso ofrece apoyo gráfico para visualizar propiedades básicas de los ítems y aspectos de fiabilidad de la prueba, mientras el segundo muestra el spreadplot asociado al Análisis Factorial Exploratorio. ViSta-CITA incluye otras siete visualizaciones de este tipo para diferentes tipos de análisis psicométricos. En nuestra experiencia, este tipo de representaciones gráficas,

utilizadas en combinación con las salidas numéricas convencionales, ayudan a lograr una mejor comprensión de los análisis y de los conceptos psicométricos involucrados. Los gráficos resultan en muchos casos más esclarecedores y generan mejor *insight* al momento de captar el significado de la información.

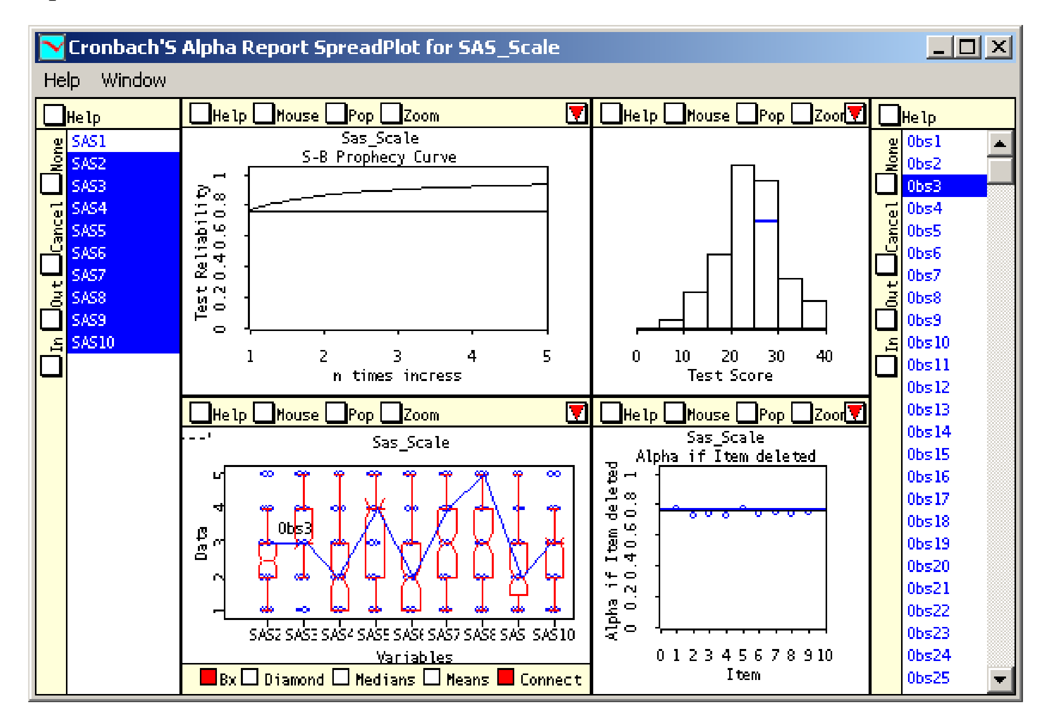


Figura 3. Imagen de spreadplot especializado del programa *ViSta-CITA*

7. Discusión

El software estadístico resulta un elemento clave para el desarrollo de cursos de análisis de datos en carreras de ciencias. Además de comprender conceptos estadísticos, los estudiantes necesitan adquirir habilidades para procesar, analizar y presentar resultados de investigación mediante la asistencia de software adecuado. En este sentido, el software no solo cumple una función meramente asistencial, pues ciertas tareas y tipos de análisis son directamente inconcebibles sin el apoyo de un programa adecuado. Por ejemplo, si pensamos en la enseñanza de los métodos de simulación basados en cómputo intensivo (e.g., bootstrap, métodos basados en permutaciones, etc.), no se puede prescindir de una herramienta informática.

El docente debe considerar varias cuestiones al momento de elegir un software apropiado (objetivos de enseñanza, contenidos que pretende trasmitir, formación del estudiante en materia informática, etc.). Entre los aspectos a considerar, la accesibilidad es un factor crítico, pues la adquisición de software profesional puede resultar muy costosa. Además, el uso de software comercial no garantiza que podamos cubrir todas las necesidades u objetivos de enseñanza planteados. Frente a las limitaciones de estos productos, los programas gratuitos como ViSta, EpiInfo u OS4 pueden proporcionar una alternativa viable en el contexto de la docencia. Lamentablemente, estos recursos tienen escasa difusión o incluso son prácticamente desconocidos por los docentes.

En este trabajo hemos enfatizado las posibilidades de un sistema gratuito en particular: el programa ViSta "The Visual Statistics System". ViSta es una herramienta que hemos utilizado en nuestra experiencia profesional con fines de desarrollo, pero creemos que es potencialmente útil en el ámbito de la enseñanza. Este sistema destaca

por su flexibilidad, simplicidad de uso y capacidades en materia gráfica, propiedades que resultan beneficiosas cuando se trata de cursos orientados a estudiantes de ciencias humanas y sociales. Asimismo, ViSta puede ser un entorno poderoso para trabajar en cursos de formación más avanzados, por ejemplo, aquellos directamente orientados a temas de programación estadística. Como hemos indicado, este sistema está desarrollado completamente en LispStat, con todas las posibilidades que este lenguaje ofrece desde el punto de vista de la programación.

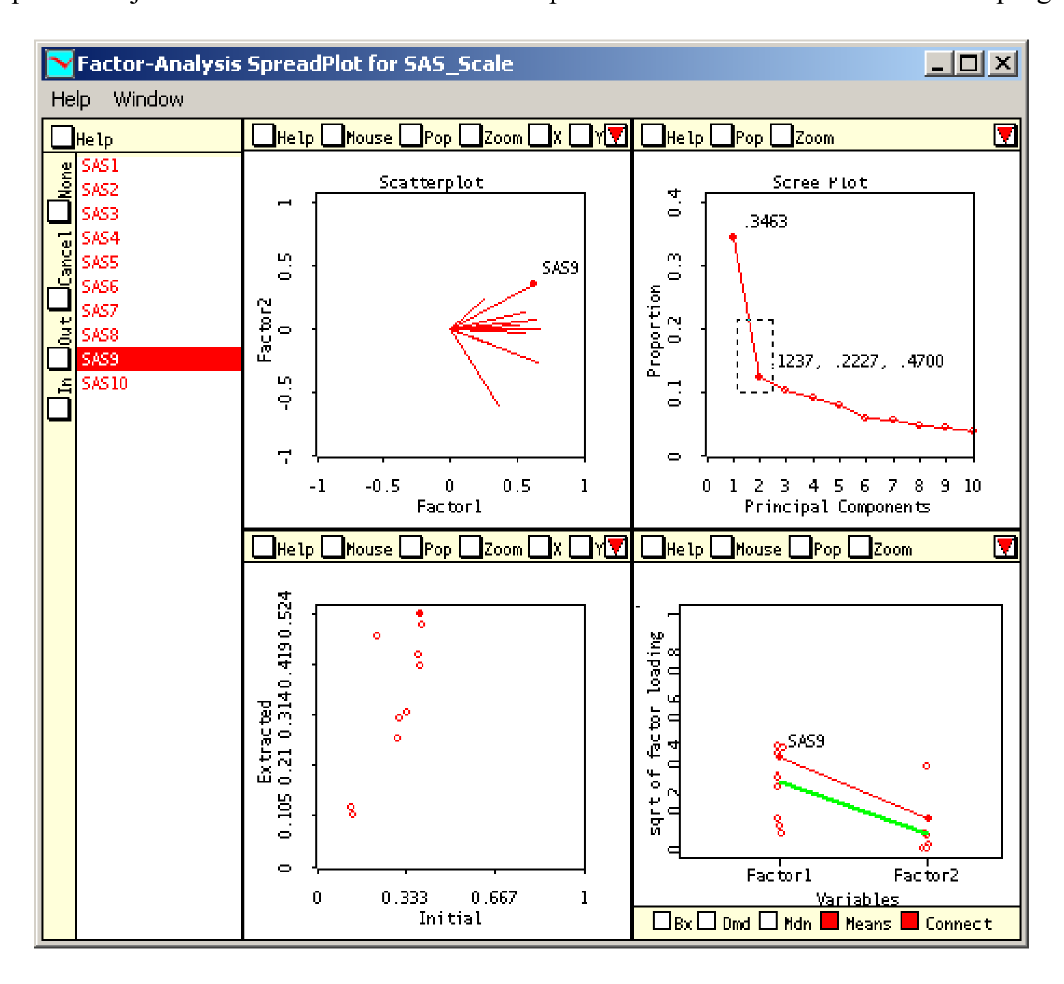


Figura 4. Imagen de spreadplot especializado del programa *ViSta-CITA*

Como limitaciones actuales de ViSta, debemos mencionar que sus versiones más recientes sólo se encuentran disponibles en inglés y para el sistema operativo Windows. Además, la documentación de usuario es algo incompleta y fragmentada. Por las características del proyecto, las posibilidades de superar estas limitaciones y mejorar las prestaciones del sistema dependen del interés de la comunidad de usuarios y del incremento en el número de desarrolladores abocados a la tarea. Esperamos que este trabajo aliente a los expertos en este sentido y contribuya a interesar a los docentes y estudiantes de estadística aplicada.

Agradecimientos

Este trabajo fue realizado con el apoyo financiero de la Universidad Nacional de Mar del Plata (Argentina), el Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina) y la Universidad de Valencia (España).

Nota

Un avance de este trabajo fue presentado en el

XXXV Coloquio Argentino de Estadística, Mar del Plata, 22-24 de Octubre, 2007.

Referencias

CDC-Centers for Disease Control And Prevention (2000). *EpiInfo 2000*. Manual de usuario. Atlanta: CDC.

Cook, R. D., y Weisberg, S. (1999). *Applied regression including computing and graphics*. New York: John Wiley & Sons.

Ledesma, R.D. (2004). Sistemas estadísticos de propósitos múltiples: una revisión de programas gratuitos. *Metodología de Encuestas, 6* (2), 105-117

Ledesma, R.D., y Molina, J.G. (2009). Classical Item and Test Analysis with Graphics: the ViSta-CITA Program. *Behavior Research Methods, 41* (4), 1161-1168.

Ledesma, R., Molina, J.G., Valero-Mora, P., y Poó (2007). Software estadístico de libre acceso en Psicología. Una librería de módulos para el sistema ViSta. *Evaluar, 7*, 19-33.

Miller, W. G. (2003). *OpenStat 3* [programa informático] [en línea] Disponible: URL: http://www.statpages.org/miller/openstat/OS3.html

R Development Core Team (2002). *R1.7.1* [programa informático] [en línea], Disponible: URL: http://www.r-project.org/

SPSS Inc. (2003) *SPSS 12.0 Base user's guide*. NJ: Prentice Hall

StataCorp LP (2009) Stata 11. Base Reference Manual. TX: Stata Corporation, University Dr East.

Statsoft Inc. (2005). Statistica Base [programa informático]. Disponible: http://www.statsoft.com

Tierney, L. (1990) Lisp-Stat An object-oriented environment for statistical computing and dynamic graphics. NY: John Wiley & Sons.

UNESCO (2002). *WinIDAMS Reference Manual*. 6ta. edición. París: The United Nations Educational, Scientific and Cultural Organization.

Young, F. W. (1996) *ViSta: The Visual Statistics System*. UNC L.L. Thurstone Psychometric Laboratory, Research Memorandum 94-1, 1996.

Young, F. W., Valero-Mora, P., y Friendly, M. (2006). *Visual Statistics. Seeing data with dynamic interactive graphics*. NJ: Wiley & Sons.

Young, F. W., Valero-Mora, P., Faldowski, R. y Bann, C. M. (2003). Gossip: The arquitecture of spreadplots. *Journal of Computational and Graphical Statistics, 12,* 80–100.